



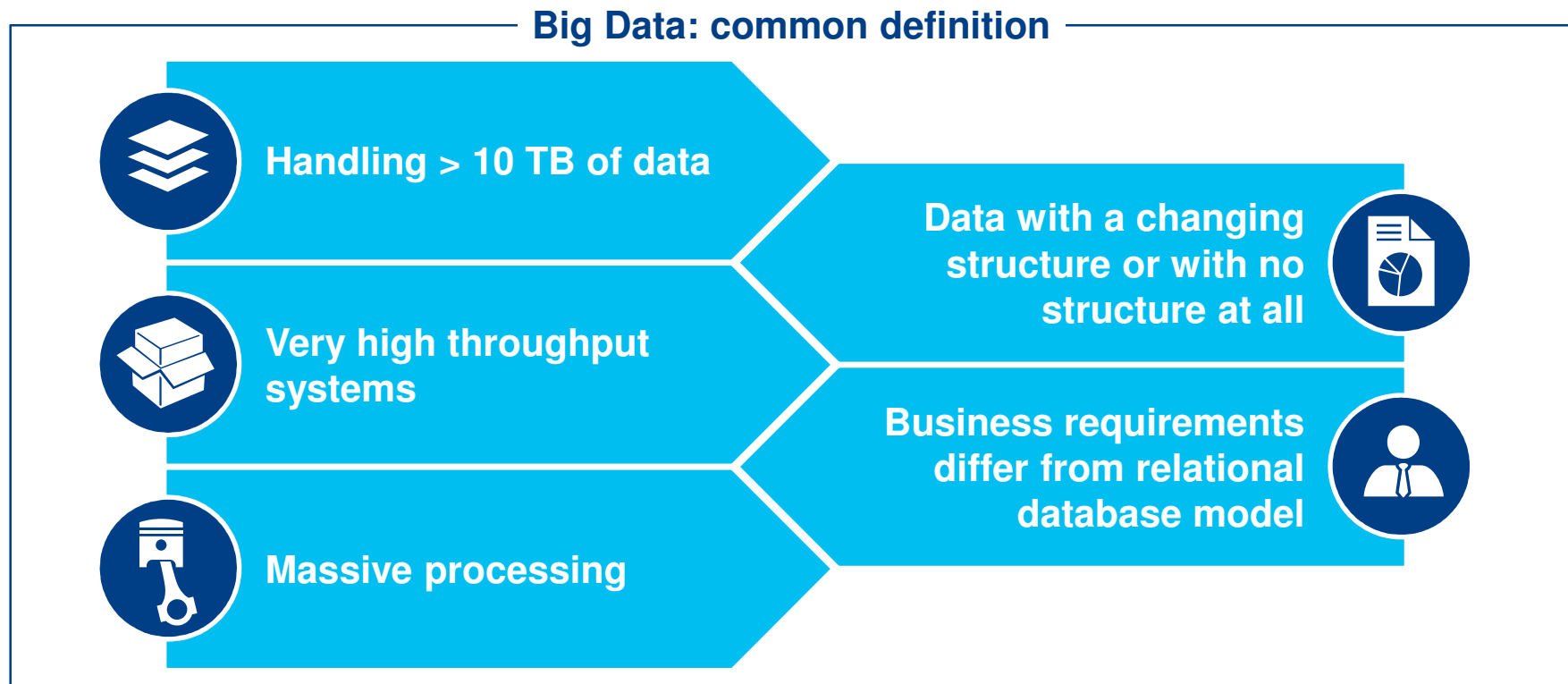
# Big Data

## Challenges and Success Factors

Deloitte Analytics  
Your data, inside out

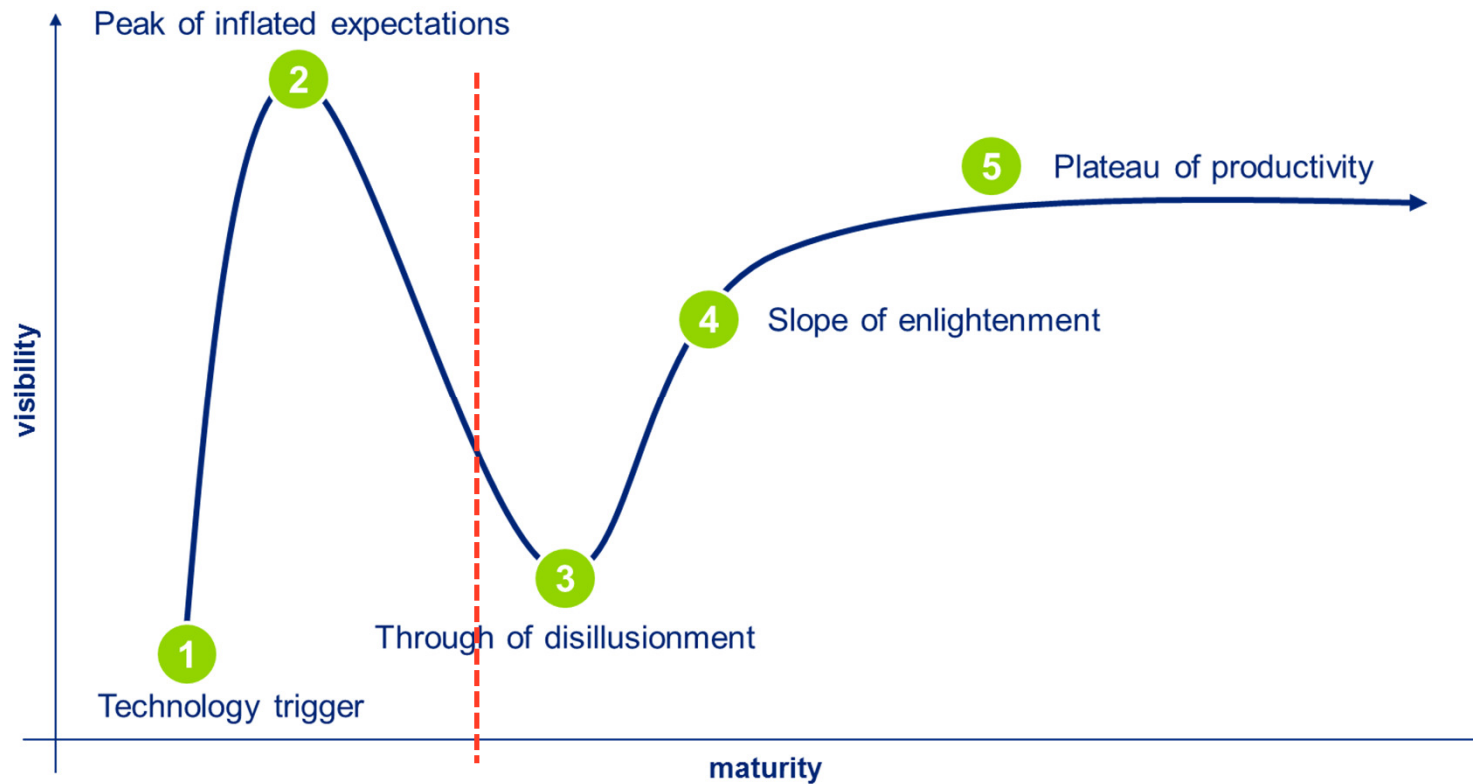


**Big Data refers to the set of problems – and subsequent technologies developed to solve them – that are hard or expensive to solve in traditional relational databases**



... However, **there is no single or agreed definition** as well as ***each Enterprise is on a different maturity level*** in the potential Big Data journey

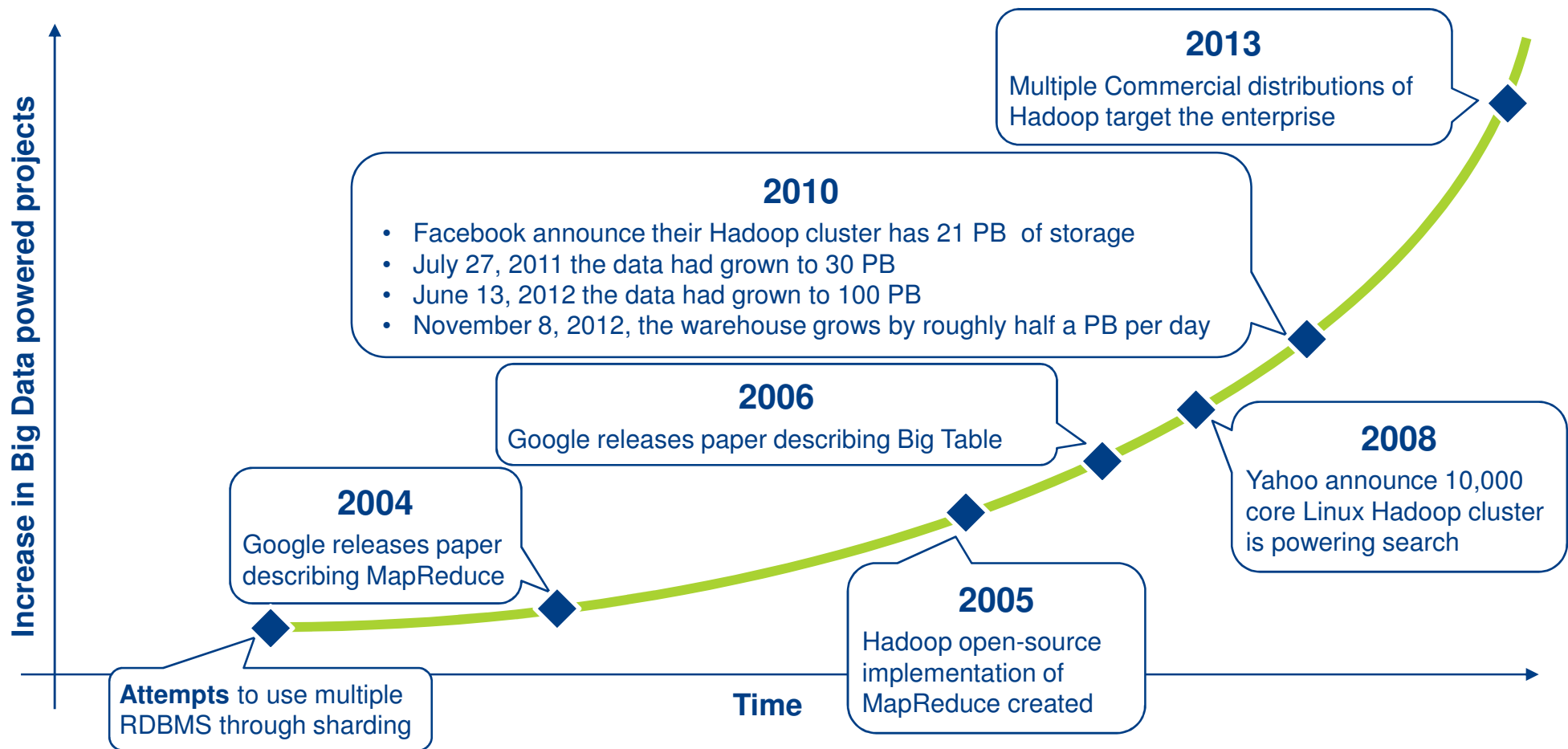
## Today positioning in the Hype-Cycle is affected by the excess of marketing messages coming both from true players and from illusionist players



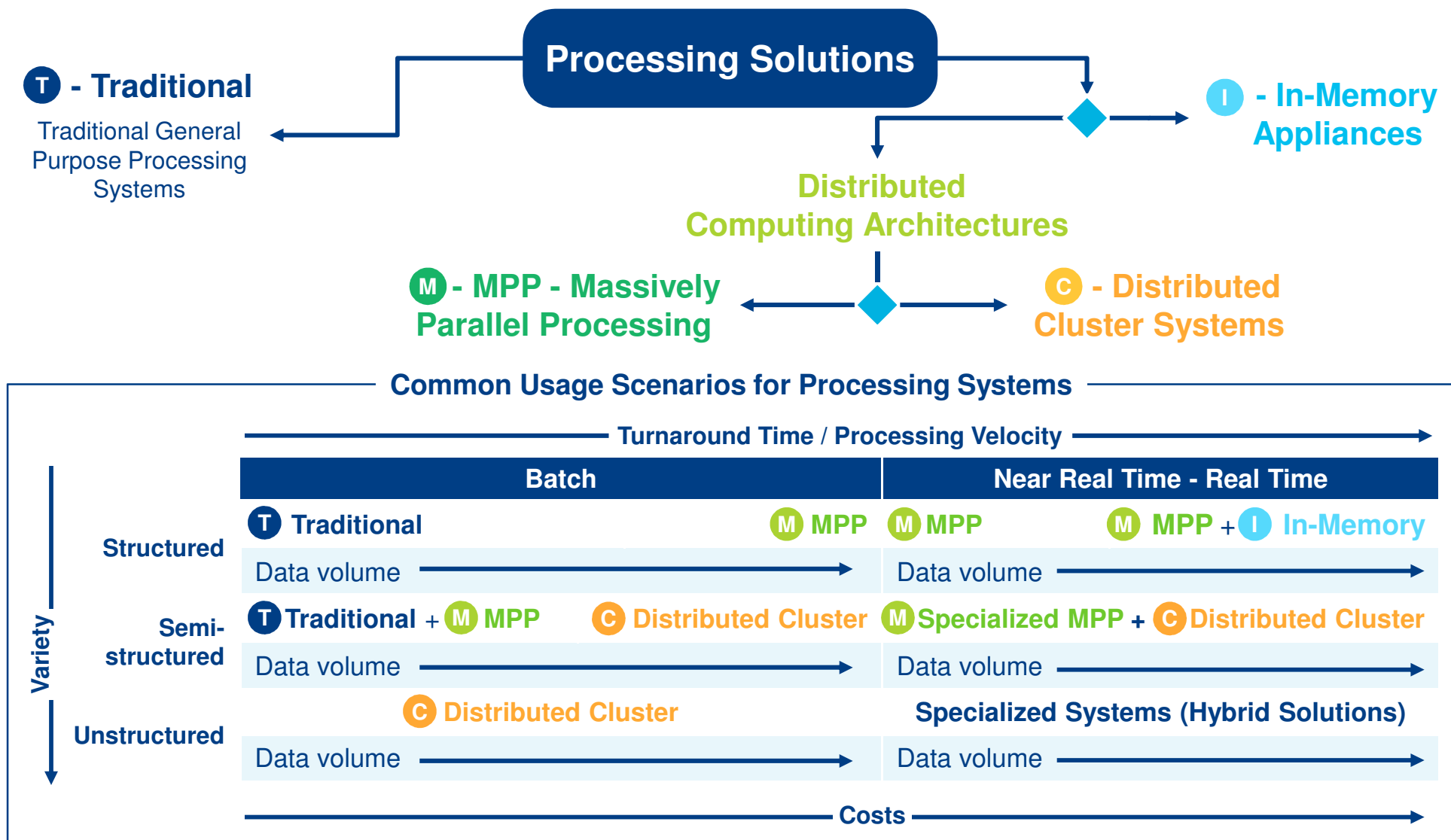
- Overused marketing term, with solutions brought in as a plug&play panacea
- Over-hyped, with few actual client references in common business world
- Buzz concentrated on social media websites / search engines

## Big Data is not just a marketing term: it is reality with a solid story and evolutionary path. Just the adoption for common business is not mature yet

- **Flexibility** of Big Data technologies is **traded** with **consistency and integrity** of RDBMS technologies
- Big Data technologies are **complementary and not a replacement** for RDBMS technologies



The challenge starts with processing: Big Data solutions can be classified based on the expected service levels being addressed and the type of underlying data...



... and continues with analysis: from the analysis standpoint, Big Data can be approached with several applications specialized for specific needs



### Search & Analysis

- **Syntactic Analysis and text mining**, with statistical models for keywords and key topics detection
- **Semantic Analysis with a Natural Language Processing engine** and ontologies, to map concepts in a specific context



### Data Visualization

- **Visual representation of Data** to communicate information clearly and effectively through graphical means for an immediate 'capture'



### Data Discovery

- **Dynamic and easy-to-use reports for freely navigating across data** with no predefined paths



### Data Mash-Up

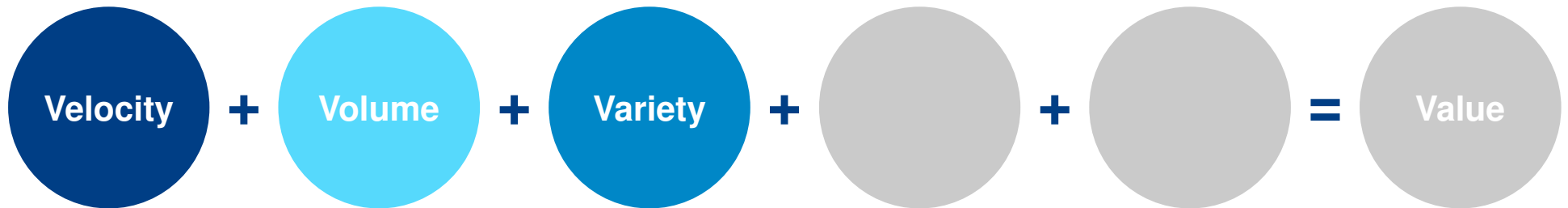
- Tools with the **ability to combine structured and unstructured data from disparate systems** and automatically organize information for search, discovery, and analysis



### Traditional Reporting

- **Static reporting for standardized access** to institutional and predefined information

**Big Data is often described by the 3 V's: Velocity, Volume and Variety: each V represents a hard problem for traditional databases**



**Velocity:** Frequency of generation is too high to be managed traditionally

48

**Hours of video**  
uploaded every  
minute to Youtube

2M

**queries on**  
**Google** every  
minute

47K

**App downloads**  
per minute via  
iTunes

**Volume:** The growth of world data is exponential

2.5

**Zettabytes of**  
world data in  
**2010**

8

**Zettabytes of**  
world data in  
**2015**

35

**Zettabytes of**  
world data in  
**2020**

**Variety:** Big Data can be structured and unstructured



**Web / Social Media**



**Machine to Machine**



**Big Transaction Data**

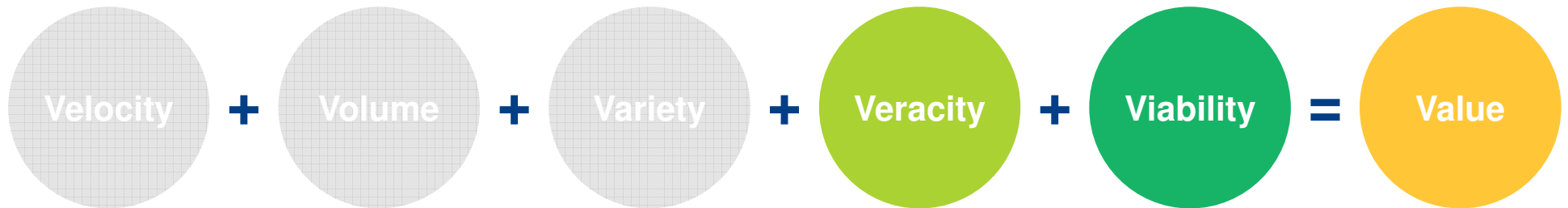


**Biometric**



**Human Generated**

However, additional V's are being proposed, to generate greater value: as the world of data grows, so does the challenge.



### Veracity: Establishing trust in data



One Third of Business leaders **do not trust the information** they use



Uncertainty is due to **inconsistency, ambiguity, latency and approximation**

### Viability: Relevance and Feasibility



**Hypothesis** - validation to determine **if the data will have a meaningful impact**



**Long Term rewards and better outcomes** from hidden relationships in data

### Value: Measuring return on investments



**Costs** – there is a serious risk of simply creating Big Costs without creating strong value



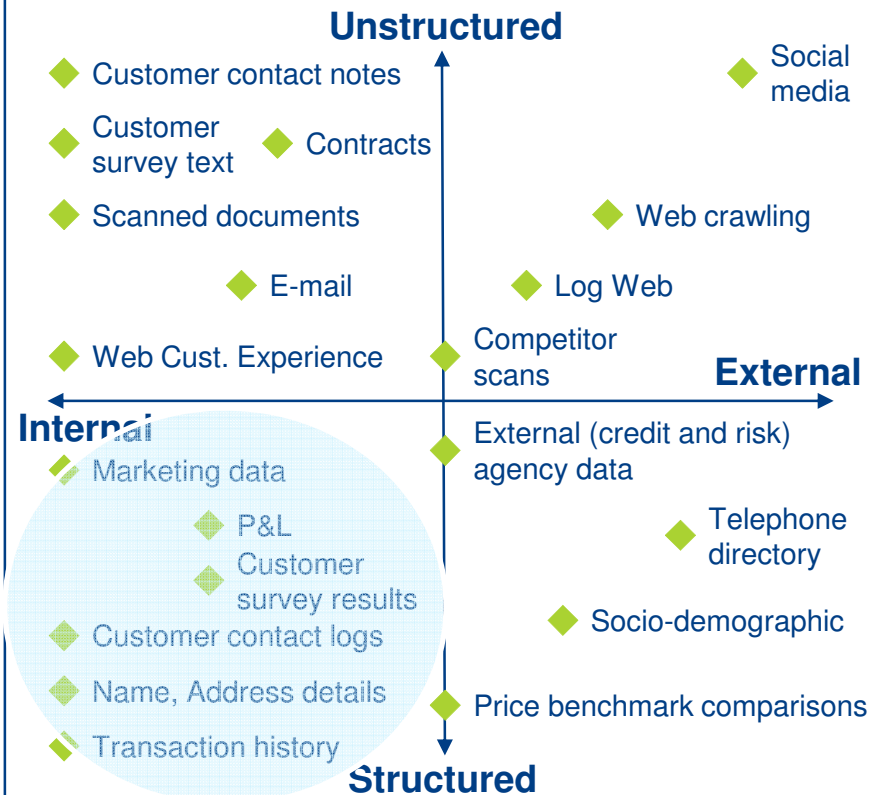
**Insights** – Sophisticated queries, counterintuitive insights and unique learning



# Big Data can enhance customer view exploiting the potential of hidden meanings

## More data sources

- Flexibility of Big Data technologies allows the usage of both
  - Internal and external data
  - Structured and unstructured data



## More insights

- Big Data can provide a **whole new set of information**, in order to reach an omni-comprehensive and multi-level customer view

### Interaction Data

- Email
- Chat transcription
- Call Center notes
- Web analytics
- In person dialogues

### Attitudinal Data

- Options
- Preferences
- Needs and Desires
- Market Research
- Social Media

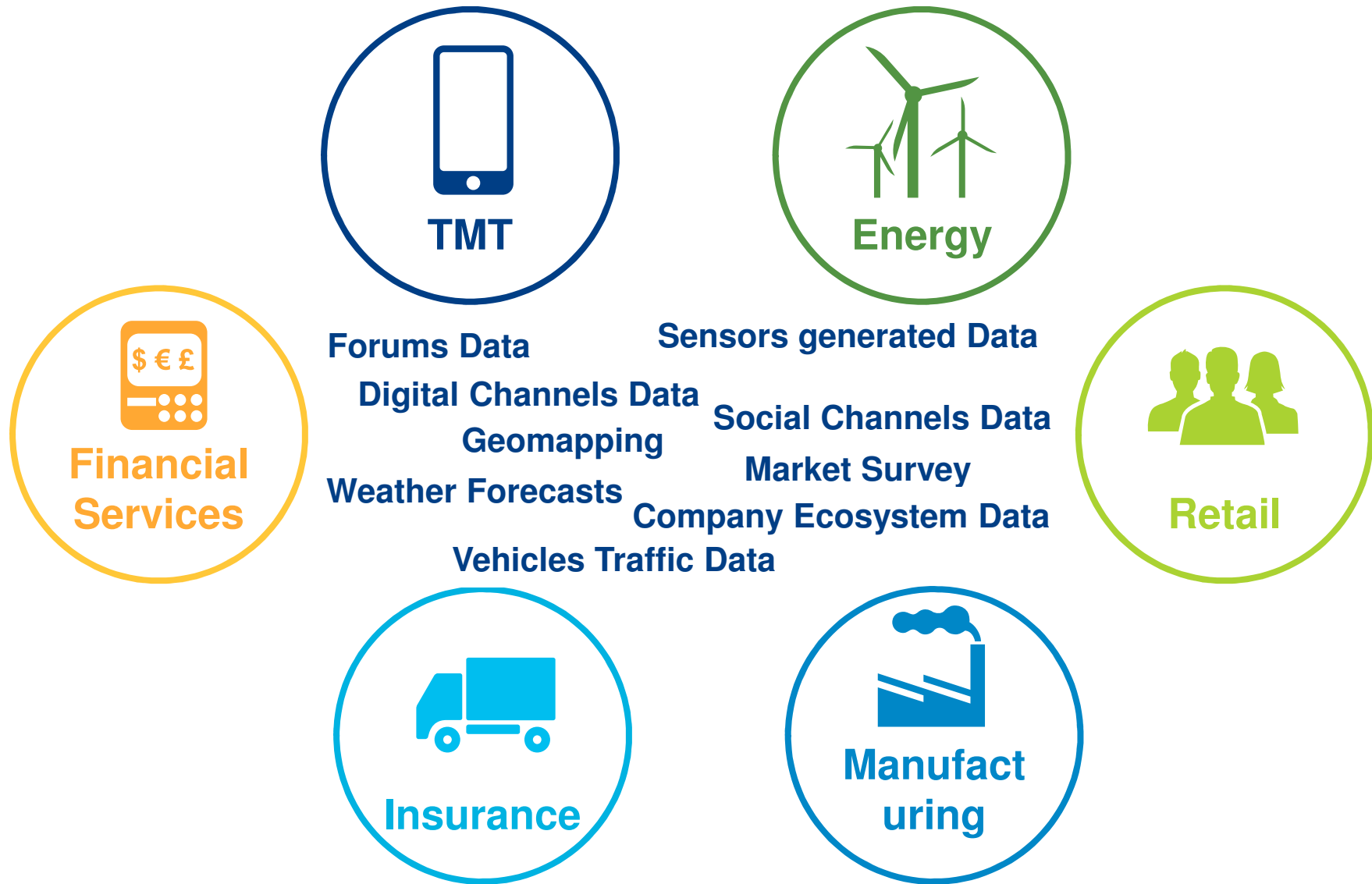
### Behavioural Data

- Orders
- Transactions
- Payment History
- Usage History

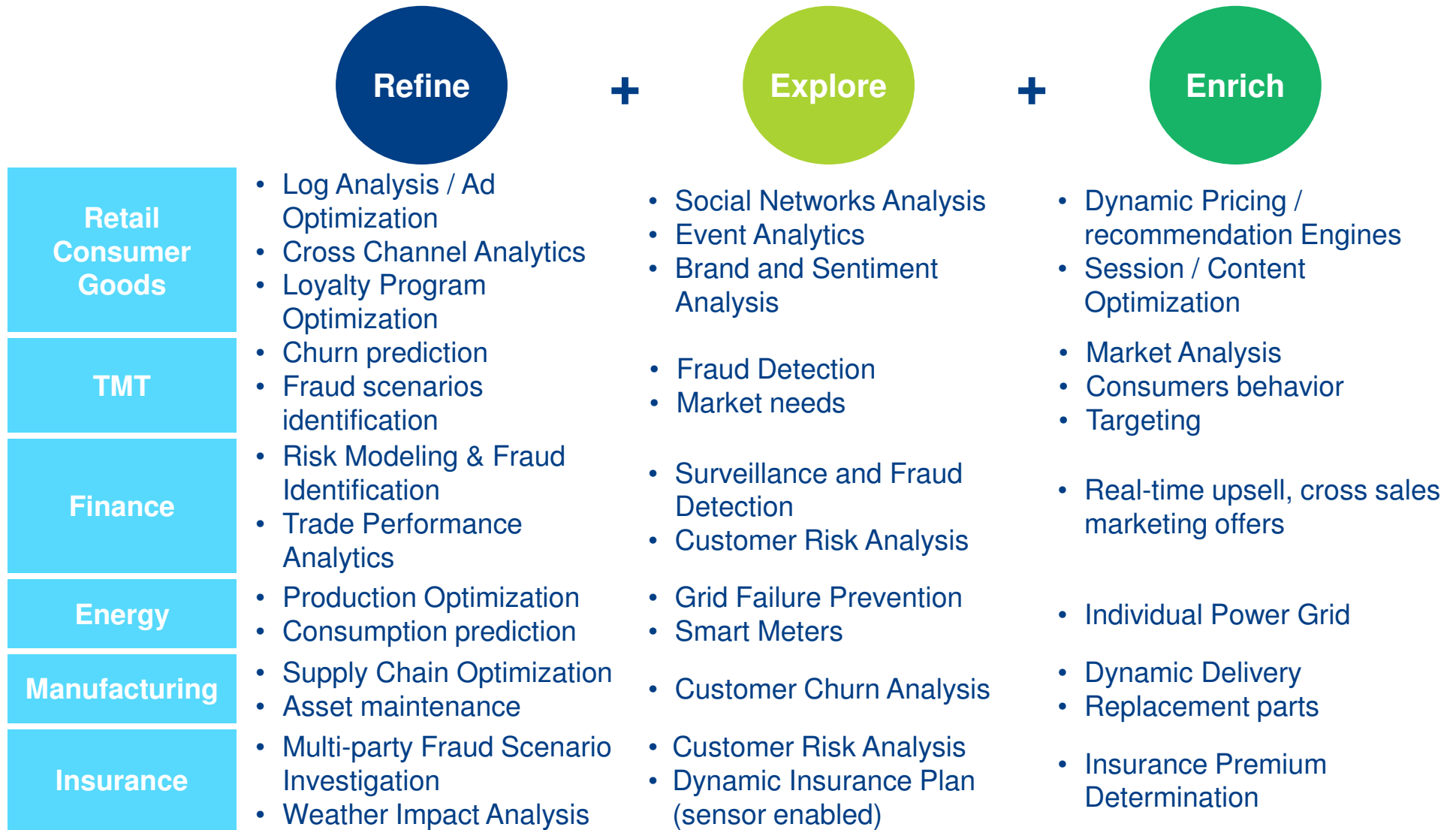
### Descriptive Data

- Attributes
- Characteristics
- Self Declared Info
- Social Geo / Demographics info

**The power of Big Data extends further away from a Social Media centric view: the following industries have already gathered scenarios requiring Big Data solution**



## For each industries is possible to evaluate enhancements based on Refinement, Exploration and Enrichment of existing scenarios



## Big Data comes with lots of challenges: Big Data provides opportunities however there are challenges that need to be addressed and overcome 1/2



### Strategy

- **Determine a strategy** how to leverage on the benefits of Big Data
- **Determine business drivers** and if Big Data can play a role in better insight
- Define **criteria** for evaluating return on investments



### Talent

- **Identify and acquire the skill sets** required to understand and leverage Big Data to add value
- Acquire **Data Scientists**, with expertise on math, statistics, data engineering, pattern recognition, advanced computing, visualization and modeling
- Organize **business analysts** team with strong knowledge of company ecosystem

## Big Data comes with lots of challenges: Big Data provides opportunities however there are challenges that need to be addressed and overcome 2/2



### Scalability

### Integration

### Deployment

### Analytics

- **Flexibility of infrastructure** to interact with extreme volume / variety of data formats
- **Cost and effort associated** with scalability
- **Increasing data volume, variety, and complexity results in increased time and investments to remove barriers** to compiling, managing and leveraging data across multiple platforms /systems
- Identifying **the best software and hardware solutions** and determining the best overall infrastructure solution; internally, externally or using a combination
- **Transitioning** from legacy systems to newer technology
- Considerable **time and money invested to create algorithms** that scale to big data volume and variety and improve user experience



### Data Quality

### Governance

### Privacy

- **Compromise of quality** due to volume and variety of data
- **Cost of maintaining all** data quality dimensions: Completeness, Validity, Integrity, Consistency, Timeliness, and Accuracy
- Identifying **relevant data protection requirements** and developing an **appropriate governance** strategy
- **Reevaluation of internal and external data policies** and regulatory environment
- **Privacy issues** related to direct and indirect use of big data sources
- Evolving **security implications** of big data

These challenges require a strong roadmap, which begins with decision makers and their crunchy questions, and proceeds to data sources and technologies

### 3 - Determine data sources

Assess:

- **Data and application landscape** including archives
- **Analytics and BI capabilities** including skills
- **Assess new technology adoptions**
- IT strategy, priorities, policies, budget and investments
- Current projects
- Current data, analytics and BI problems

### 4 - Identify / Define Use Cases

Based on the assessments and business priorities  
**identify and prioritize big data use cases**

### 2 - Identify Opportunities

Brainstorm and **ask crunchy questions**

### 1 - Strategic plan

Identify **strategic priorities**

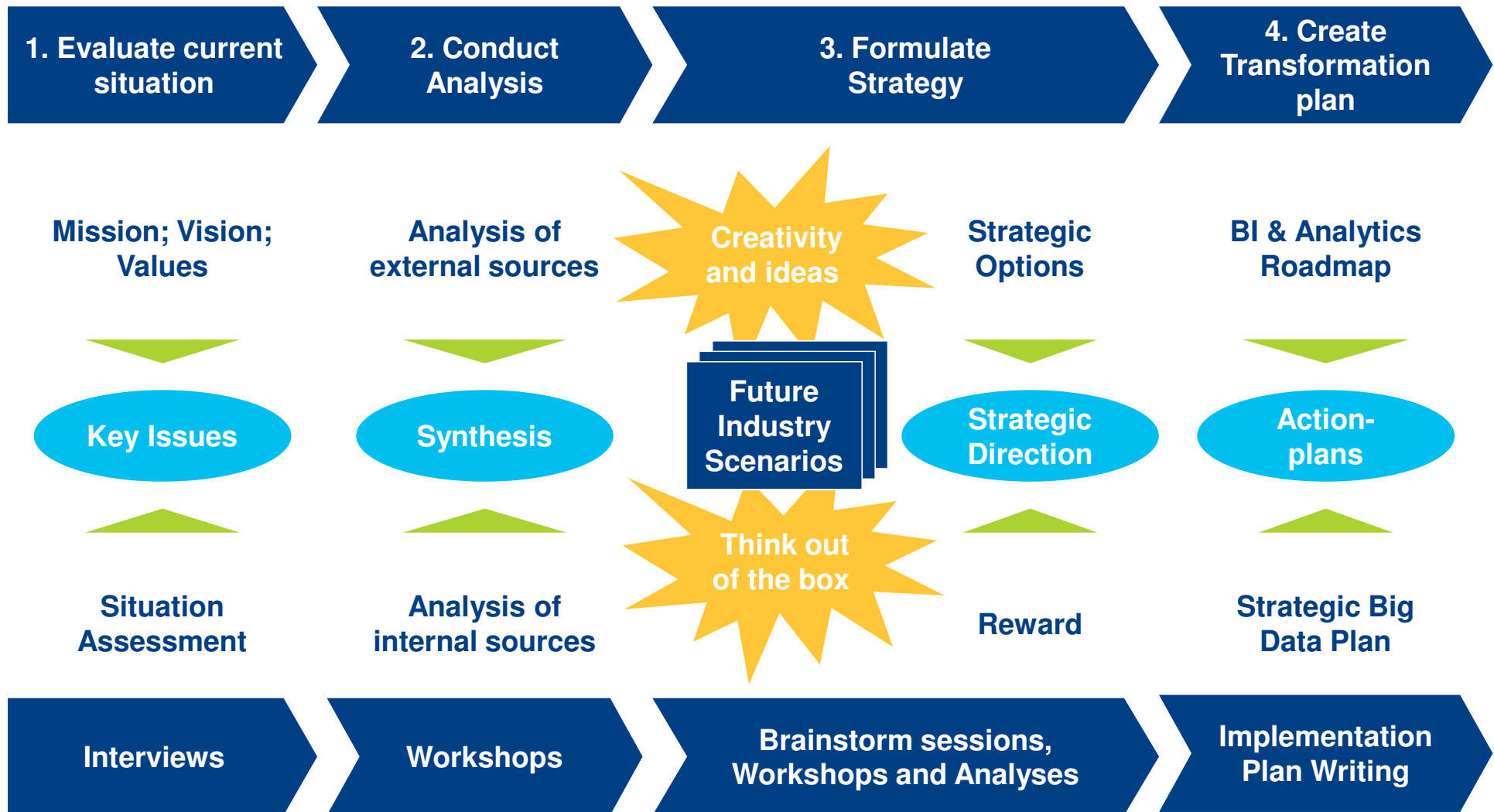
### 5 - Pilots and Prototypes

Identify **tools, technologies and processes** for use cases and implement pilots and prototypes

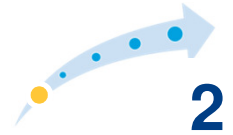
### 6 - Adopt in Production

**Prioritize and implement successful, high value initiatives** in production

Every Big Data project starts with a short planning and scoping phase...



... and goes on identifying strategic opportunities asking crunchy questions for “sticky” business issues



### Customers and social media

- What's the **buzz about your company online**, and how could it impact sales?
- What are **analysts saying about your organization**? What about **customers and online influencers**?
- Who are the **next 1,000 customers you'll lose** - and why?
- Which trade promotion programs have the **highest impact on profitability**?
- What **factors most influence customer loyalty**? Why?
- How do **factors such as politics and demographics affect the price** your customers are willing to pay?
- Which factors have the **most adverse effects on customer satisfaction**?



### Sustainability and supply chain

- Which facilities are **using more energy than they should**?
- Which suppliers are **at risk of going out of business**?
- What is the **impact of shipping costs** on pricing?
- Which locations offer the **best options** for setting up your next distribution center?

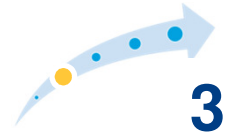


### Employees and risk

- Which **new-hire characteristics best reflect your organization's** risk intelligence profile?
- Which are **most likely to steal** from you?
- Why do **high-potential employees leave your company**? What would cause them to stay?

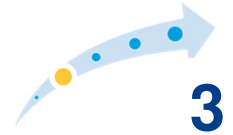


## Bringing Big Data into the current Business Ecosystem leads to a multitude of difficult questions to be answered (1/2)



- What **data sources should be collected** and how can they be acquired efficiently?
- Should **retention** be provided for those data?
- How **intensively** will those data be processed?
- **How is data quality managed across so many sources of data**, many of which come from outside the organization, such as public social networks?
- **What structure can be derived from non-traditional data sources** (documents, Web logs, video streams, etc.) to make storage, analysis, and ultimately decision-making easier?
- **How can non-traditional unstructured data be integrated** with data stored in traditional transactional systems?
- **How can decision-makers comprehend the results** of analyzing so much data quickly enough to act?
- **What data governance is appropriate** when analysis is distributed, needs change and data definitions and schemas evolve over time?
- **What architectures and algorithms can be used** to decompose problems and data for rapid execution in parallel environments?

## Bringing Big Data into the current Business Ecosystem leads to a multitude of difficult questions to be answered (2/2)



- What levels of **availability and reliability** are possible in mission-critical applications, as data volumes are so large?
- Is **specialized hardware required for a particular need**, or can low-cost commodity hardware be leveraged to scale processing?
- Given the specialized nature of processing needed, **is cloud computing an appropriate platform choice**, and if so, what variant of cloud computing (public, private, hybrid) is needed?
- How can **security and privacy concerns be factored into the design of a Big Data environment** to reduce vulnerability to external and internal threats?
- How are **regulations around audit trails and data destruction to be interpreted** in a Big Data environment?
- **What intellectual property, licensing, and data protection considerations apply** when Big Data environments are distributed across organizational and national boundaries?
- **How can current IT skill sets best be leveraged in evolving the infrastructure** to include Big Data?

## Approaching correctly to all suggestions shown will allow avoiding common pitfalls

- ✓ **Do not approach it as a new technology trend: it's about different trends coming together**  
(new technologies; new data / new domains; new analysis paradigms)
- ✓ **Use technologies with awareness**
- ✓ **Don't trust data just because they exist. Be selective.**
- ✓ **Big Data is not mandatory. Adopt it if you can really gain advantages**
- ✓ **Enforce collaboration among Enterprise Business Units**
- ✓ **Do not approach it as an IT topic –**  
Big Data fails without a strong interlock with Business:
  - Improved Customer Engagement
  - Dynamic Provisioning
  - Near Real Time decision process
- ✓ **Re-Arrange BI Domain organization**
  - Traditional organizational model will be discontinued in few years, as new roles are emerging
  - Traditional BI implementation lifecycle reactivity is going to impose new models
- ✓ **Do not lose focus on traditional BI**
  - Data quality and Data Governance issues can be amplified with Data Fusion through Big Data
  - Strong Data Fusion between structured and unstructured data is the Key Success factor

**Deloitte.**